

(19)日本国特許庁(JP)

(12) 公 開 特 許 公 報(A)

(11)特許出願公開番号

特開2023-179189
(P2023-179189A)

(43)公開日 令和5年12月19日(2023. 12. 19)

(51)Int. Cl.	F I	テーマコード (参考)
<i>G 1 0 L 25/60 (2013. 01)</i>	G 1 0 L 25/60	
<i>G 1 0 L 25/06 (2013. 01)</i>	G 1 0 L 25/06	
<i>G 1 0 L 25/18 (2013. 01)</i>	G 1 0 L 25/18	

審査請求 未請求 請求項の数 14 O L (全 18 頁)

(21)出願番号	特願2022-92345(P2022-92345)	(71)出願人	504145283 国立大学法人 和歌山大学 和歌山県和歌山市栄谷 9 3 0 番地
(22)出願日	令和4年6月7日(2022. 6. 7)	(74)代理人	100111567 弁理士 坂本 寛
		(72)発明者	入野 俊夫 和歌山県和歌山市栄谷 9 3 0 番地 国立大 学法人和歌山大学内

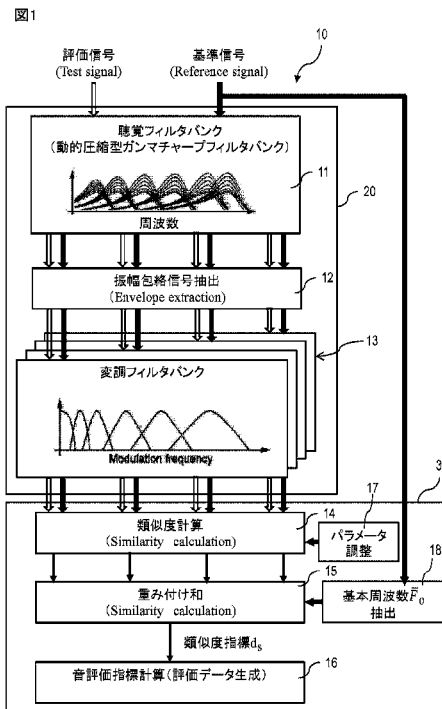
(54)【発明の名称】音の評価指標計算方法、評価データを生成する方法、音の評価装置、及びコンピュータプログラム

(57)【要約】

【課題】評価信号と基準信号との音圧レベルを無くすための正規化をしなくても、音の客観的な評価指標を求めることを可能にする技術を提供する。

【解決手段】開示の評価指標計算方法は、音の評価信号と基準信号とを周波数分析して、前記評価信号の第1特徴量と前記基準信号の第2特徴量とを求め、前記第1特徴量と第2特徴量との類似度を求め、前記類似度に基づいて、前記評価信号の評価指標を計算する、ことを備え、前記類似度は、前記第1特徴量及び前記第2特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められる。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

音の評価信号と基準信号とを周波数分析して、前記評価信号の第 1 特徴量と前記基準信号の第 2 特徴量とを求め、

前記第 1 特徴量と第 2 特徴量との類似度を求め、

前記類似度に基づいて、前記評価信号の評価指標を計算する、

ことを備え、

前記類似度は、前記第 1 特徴量及び前記第 2 特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第 1 特徴量及び重み付き第 2 特徴量を用いて求められる

10

音の評価指標計算方法。

【請求項 2】

前記評価信号と前記基準信号とを周波数分析することは、前記評価信号と前記基準信号とが入力信号として与えられる分析器によって行われ、

前記分析器は、周波数分析のための 1 又は複数のフィルタバンクを備える

請求項 1 に記載の評価指標計算方法。

【請求項 3】

前記類似度は、前記重み付き第 1 特徴量及び前記重み付き第 2 特徴量を用いて計算されるコサイン類似度である

請求項 1 に記載の評価指標計算方法。

20

【請求項 4】

前記調整可能パラメータは、

前記評価信号の音響特性、

前記評価信号を受聴する聴取者の受聴特性、及び

前記評価信号の受聴環境特性

からなる群から選択されるいずれか一つの特性又はこれらの 2 以上の特性の組み合わせによって影響を受ける特性に基づいて調整される

請求項 1 に記載の評価指標計算方法。

【請求項 5】

前記調整可能パラメータは、前記評価信号を受聴する聴取者に対する聴覚検査の結果に基づいて調整される

請求項 1 に記載の評価指標計算方法。

30

【請求項 6】

前記評価信号と前記基準信号とを周波数分析することは、周波数分析のための複数の周波数チャンネル毎に前記第 1 特徴量及び前記第 2 特徴量を求めることを備え、

前記類似度は、前記複数の周波数チャンネル毎に求められ、

前記評価指標は、前記複数の周波数チャンネルにおける前記類似度を統合した類似度指標に基づいて計算される

請求項 1 に記載の評価指標計算方法。

【請求項 7】

前記類似度指標は、前記複数の周波数チャンネルにおける前記類似度それぞれを重み付け計算することで求められる

請求項 6 に記載の評価指標計算方法。

40

【請求項 8】

前記複数の周波数チャンネルにおける前記類似度それぞれを重み付け計算するための重みは、前記基準信号の基本周波数に基づいて決定される

請求項 7 に記載の評価指標計算方法。

【請求項 9】

前記 1 又は複数のフィルタバンクは、動的圧縮型ガンマチャープフィルタバンクを含む
請求項 2 に記載の音の評価指標計算方法。

50

【請求項 1 0】

前記 1 又は複数のフィルタバンクは、聴覚系のモデルである聴覚フィルタバンクを含む請求項 2 に記載の音の評価指標計算方法。

【請求項 1 1】

前記第 1 特徴量及び前記第 2 特徴量は、前記聴覚フィルタバンクの出力の振幅包絡に基づいて求められる

請求項 1 0 に記載の音の評価指標計算方法。

【請求項 1 2】

評価データを生成する方法であって、

音の評価信号と基準信号とを周波数分析して、前記評価信号の第 1 特徴量と前記基準信号の第 2 特徴量とを求め、

前記第 1 特徴量と第 2 特徴量との類似度を求め、

前記類似度に基づいて、前記評価信号の評価データを生成する、

ことを備え、

前記類似度は、前記第 1 特徴量及び前記第 2 特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第 1 特徴量及び重み付き第 2 特徴量を用いて求められる

評価データを生成する方法。

【請求項 1 3】

音の基準信号と評価信号とが入力信号として与えられる分析器と、

前記評価信号に対する評価データを生成する評価器と、

を備え、

前記分析器は、前記入力信号として与えられた前記基準信号と前記評価信号とを周波数分析して、前記基準信号の第 1 特徴量と前記評価信号の第 2 特徴量とを出力するよう構成され、

前記評価器は、前記第 1 特徴量と前記第 2 特徴量との類似度を求め、前記類似度に基づいて、前記評価信号に対する評価データを生成するよう構成され、

前記類似度は、前記第 1 特徴量及び前記第 2 特徴量を重み付けするための調整パラメータによって重み付けられた重み付き第 1 特徴量及び重み付き第 2 特徴量を用いて求められる

音の評価装置。

【請求項 1 4】

音の評価信号と基準信号とを周波数分析して、前記評価信号の第 1 特徴量と前記基準信号の第 2 特徴量とを求め、

前記第 1 特徴量と第 2 特徴量との類似度を求め、

前記類似度に基づいて、前記評価信号の評価指標を計算する、

ことを備える処理を、コンピュータに実行させるためのコンピュータプログラムであって、

前記類似度は、前記第 1 特徴量及び前記第 2 特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第 1 特徴量及び重み付き第 2 特徴量を用いて求められる、

コンピュータプログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本開示は、音の評価指標計算方法、評価データを生成する方法、音の評価装置、及びコンピュータプログラムに関する。

【背景技術】**【0002】**

特許文献 1 は、音声明瞭度計算方法を開示している。特許文献 1 に開示の技術では、複

10

20

30

40

50

数のフィルタバンクを用いて、入力されたクリーン音声（基準信号）の特徴量である時間的な振幅包絡信号と強調音声（評価信号）の特徴量である振幅包絡信号との差分である歪み成分が求められる。歪み成分を用いて、音声明瞭度が計算される。なお、強調音声は、クリーン音声に雑音が重畳された音声（雑音重畳音声）である。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】国際公開第2019/027053号

【発明の概要】

【0004】

10

特許文献1に開示の技術では、フィルタバンクに入力される評価信号（雑音重畳音声）及び基準信号（クリーン音声）が、フィルタバンクに入力される前に、評価信号及び基準信号の音圧レベル差を無くすため、両信号の正規化をする必要が生じる。

【0005】

正規化する必要があるのは、特許文献1に開示の技術では、評価信号の特徴量と基準信号の特徴量との差分を歪み成分として忠実に抽出するためである。歪み成分を忠実に抽出するためには、評価信号と基準信号の特徴量を求めて歪み成分を抽出する前に、両信号の音圧レベル差を無くす必要がある。音圧レベル差を無くすための正規化は、例えば、両信号のrms（root mean square）レベルの正規化である。

【0006】

20

仮に、評価信号と基準信号との間に音圧レベル差があるまま両信号の特徴量を求めて、それらの特徴量を差分として求めると、その差分には、両信号の特徴量の歪み成分だけでなく、評価信号と基準信号との間の音圧レベル差に起因する成分も含むことになる。この場合、歪み成分だけを忠実に抽出できていないため、音声明瞭度を精度よく計算することができない。

【0007】

以上のように、特許文献1に開示の技術では、歪み成分だけを忠実に抽出するため、フィルタバンクに入力される前の評価信号と基準信号との間の音圧レベル差を無くすための正規化が必要となる。

【0008】

30

しかし、特許文献1に開示の技術のように正規化が必要であると、評価信号と基準信号の音圧レベルが本質的に異なる場合に対処できないという問題が生じる。

【0009】

基準信号に対する音圧レベルが本質的に異なる評価信号は、例えば、補聴器信号である。補聴器信号は、補聴器から出力されて難聴者に与えられる。補聴器信号は、難聴者に与えられる信号であるため、基準信号に対して音圧レベルが大きく異なるという特性を本質的に有する。

【0010】

したがって、補聴器信号と基準信号とを正規化してしまうと、両信号の音圧レベル差が失われる。音圧レベル差が失われると、基準信号に対して本質的に音圧が異なる補聴器信号としての本質的な情報が失われおそれがある。この結果、難聴者に対する補聴器信号の音声明瞭度（ここでの音声明瞭度は、例えば、難聴者の音声了解度といってもよい）などの評価指標を精度よく計算することが困難になる。

40

【0011】

したがって、評価信号と基準信号との音圧レベルを無くすための正規化をしなくても、音の客観的な評価指標を求めることを可能にする技術が望まれる。

【0012】

本開示のある側面は、音の評価指標計算方法である。開示の評価指標計算方法は、音の評価信号と基準信号とを周波数分析して、前記評価信号の第1特徴量と前記基準信号の第2特徴量とを求め、前記第1特徴量と第2特徴量との類似度を求め、前記類似度に基づい

50

て、前記評価信号の評価指標を計算する、ことを備え、前記類似度は、前記第1特徴量及び前記第2特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められる。

【0013】

本開示の他の側面は、評価データを生成する方法である。本開示のさらに他の観点は、音の評価装置である。本開示のさらに他の観点は、コンピュータプログラムである。

【0014】

更なる詳細は、後述の実施形態として説明される。

【図面の簡単な説明】

【0015】

【図1】図1は、実施形態に係る評価装置のブロック図である。

【図2】図2は、実施形態に係る評価装置における計算方法を示す図である。

【図3】図3は、防音室実験の結果を示す図である。

【図4】図4は、遠隔実験の結果を示す図である。

【図5】図5は、 $\alpha = 0.55$ であるときのGSEIによる予測結果を示す図である。

【図6】図6は、 $\alpha = 0.60$ であるときのGSEIによる予測結果を示す図である。

【図7】図7は、従来法ESTOIによる予測結果を示す図である。

【図8】図8は、従来法HASP Iによる予測結果を示す図である。

【発明を実施するための形態】

【0016】

< 1. 音の評価指標計算方法、評価データを生成する方法、音の評価装置、及びコンピュータプログラムの概要 >

【0017】

(1) 実施形態に係る評価指標計算方法は、音の評価信号と基準信号とを周波数分析して、前記評価信号の第1特徴量と前記基準信号の第2特徴量とを求め、前記第1特徴量と第2特徴量との類似度を求め、前記類似度に基づいて、前記評価信号の評価指標を計算する、ことを備える。

【0018】

評価信号は、評価指標によって評価されるべき音の信号である。評価信号は、例えば、音声、音声ではない音響信号、その他の音の信号である。基準信号は、評価の基準となる信号であり、例えば、ノイズを含まないクリーン音声の信号である。周波数分析の方法は、特に限定されないが、好ましくは、後述のようにフィルタバンクが用いられ得る。

【0019】

第1特徴量は、評価信号を周波数分析して得られる特徴量である。第2特徴量は、基準信号を周波数分析して得られる特徴量である。

【0020】

類似度は、評価信号の第1特徴量と基準信号の第2特徴量とから求められるため、評価信号が基準信号に似ていれば、類似度が高くなり、評価信号が基準信号に似ていなければ、類似度が低くなる。類似度の求め方は、特に限定されないが、好ましくは、後述のようにコサイン類似度の求め方に準じて求められる。

【0021】

実施形態において、類似度は、調整可能パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められる。このため、評価信号と基準信号との音圧レベル差があったとしても、類似度を求める際に、その音圧レベル差に応じて調整可能パラメータを調整すればよい。したがって、評価信号と基準信号との音圧レベルを無くすための正規化を予めしなくても評価指標を計算できる。

【0022】

(2) 前記評価信号と前記基準信号とを周波数分析することは、前記評価信号と前記基準信号とが入力信号として与えられる分析器によって行われるのが好ましい。前記分析器は、周波数分析のための1又は複数のフィルタバンクを備え得る。周波数分析には、1つの

10

20

30

40

50

フィルタバンクが用いられてもよいし、複数のフィルタバンクが用いられてもよい。複数のフィルタバンクは、例えば、聴覚フィルタバンク及び変調フィルタバンクを備え得る。

【0023】

(3) 前記類似度は、前記第1重み付け特徴量及び前記重み付け第2特徴量を用いて計算されるコサイン類似度であるのが好ましい。

【0024】

(4) 前記調整可能パラメータは、前記評価信号の音響特性、前記評価信号を受聴する聴取者の受聴特性、及び前記評価信号の受聴環境特性、からなる群から選択されるいずれか一つの特性又はこれらの2以上の特性の組み合わせによって影響を受ける特性に基づいて調整されるのが好ましい。

【0025】

評価信号の音響特性は、例えば、基準信号に対する評価信号の音圧レベル差である。また、音響特性は、例えば、評価信号としての補聴器信号の特性である。受聴特性は、例えば、聴覚検査によって測定され得る聴取者の特性である。受聴環境特性は、聴取者が、音を聴く際に、聴取者がおかれた環境の特性であり、ノイズとなる環境音、音を出力する機器性能等によって影響を受ける特性である。なお、環境による影響を受ける状態で聴覚検査を行うことにより、受聴特性及び受聴環境特性の組み合わせによって影響を受けた特性が測定され得る。パラメータは、これらの特性に基づいて調整され得る。

【0026】

さらに、パラメータは、音響特性と受聴特性との組み合わせによって影響を受けた特性に基づいて調整され得る。また、パラメータは、音響特性と受聴環境特性との組み合わせによって影響を受けた特性に基づいて調整され得る。さらにまた、パラメータは、音響特性と、受聴特性と、受聴環境特性との組み合わせによって影響を受けた特性に基づいて調整され得る。

【0027】

(5) 前記調整可能パラメータは、前記評価信号を受聴する聴取者に対する聴覚検査の結果に基づいて調整されてもよい。

【0028】

(6) 前記評価信号と前記基準信号とを周波数分析することは、周波数分析のための複数の周波数チャンネル毎に前記第1特徴量及び前記第2特徴量を求めることを備え、前記類似度は、前記複数の周波数チャンネル毎に求められ、前記評価指標は、前記複数の周波数チャンネルにおける前記類似度を統合した類似度指標に基づいて計算されるのが好ましい。

【0029】

(7) 前記類似度指標は、前記複数の周波数チャンネルにおける前記類似度それぞれを重み付け計算することで求められるのが好ましい。

【0030】

(8) 前記複数の周波数チャンネルにおける前記類似度それぞれを重み付け計算するための重みは、前記基準信号の基本周波数に基づいて決定されるのが好ましい。

【0031】

(9) 前記1又は複数のフィルタバンクは、動的圧縮型ガンマチャープフィルタバンクを含むのが好ましい。

【0032】

(10) 前記1又は複数のフィルタバンクは、聴覚系のモデルである聴覚フィルタバンクを含むのが好ましい。

【0033】

(11) 前記第1特徴量及び前記第2特徴量は、前記聴覚フィルタバンクの出力の振幅包絡に基づいて求められるのが好ましい。

【0034】

(12) 実施形態に係る評価データを生成する方法は、音の評価信号と基準信号とを周波数分析して、前記評価信号の第1特徴量と前記基準信号の第2特徴量とを求め、前記第1

10

20

30

40

50

特徴量と第2特徴量との類似度を求め、前記類似度に基づいて、前記評価信号の評価データを生成する、ことを備える。ここでの評価データを生成する方法は、評価データという物を生産する方法であることが意図されている。前記類似度は、前記第1特徴量及び前記第2特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められるのが好ましい。

【0035】

(13)実施形態に係る評価装置は、音の基準信号と評価信号とが入力信号として与えられる分析器と、前記評価信号に対する評価データを生成する評価器と、を備える。前記分析器は、前記入力信号として与えられた前記基準信号と前記評価信号とを周波数分析して、前記基準信号の第1特徴量と前記評価信号の第2特徴量とを出力するよう構成されているのが好ましい。前記評価器は、前記第1特徴量と前記第2特徴量との類似度を求め、前記類似度に基づいて、前記評価信号に対する評価データを生成するよう構成され、前記類似度は、前記第1特徴量及び前記第2特徴量を重み付けするための調整パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められるのが好ましい。

10

【0036】

(14)実施形態に係るコンピュータプログラムは、音の評価信号と基準信号とを周波数分析して、前記評価信号の第1特徴量と前記基準信号の第2特徴量とを求め、前記第1特徴量と第2特徴量との類似度を求め、前記類似度に基づいて、前記評価信号の評価指標を計算する、ことを備える処理を、コンピュータに実行させるためのコンピュータプログラムである。前記類似度は、前記第1特徴量及び前記第2特徴量を重み付けするための調整可能パラメータによって重み付けられた重み付き第1特徴量及び重み付き第2特徴量を用いて求められるのが好ましい。コンピュータプログラムは、記録媒体に格納され得る。記録媒体は、コンピュータ読取可能な非一時的な媒体であるのが好ましい。

20

【0037】

<2.音の評価指標計算方法、評価データを生成する方法、音の評価装置及びコンピュータプログラムの例>

【0038】

以下では、音評価指標計算方法、評価データを生成する方法、及び音の評価装置の一例として、本発明者が開発したGammachirp Envelope Similarity Index (GESI)という手法を用いた装置を、図面を参照しつつ、説明する。以下では、<2.1>においてGESIに関する背景を説明し、<2.2>においてGESIのためのシステム等について説明し、<2.3>において主観評価実験の結果を示し、<2.4>において、客観評価結果を示し、<2.5>において、GESIについてのまとめを示す。ただし、GESIは、一例であり、本発明はGESIに限定されるものではない。

30

【0039】

<2.1 GESIに関する背景>

【0040】

超高齢社会を迎えた現在、高齢難聴者個人ごとに聞こえを改善できる音声・聴覚支援機材の開発が急務である。そこでは、音声強調処理/雑音抑圧処理が必須の技術となっている。この開発のためには、聴取者による強調音声の主観評価実験と、その結果を精度良く予測する客観評価指標が不可欠である。その中で、評価対象とする信号(評価信号)に対して、基準となる信号(基準信号)の情報を用いて評価を行う、イントゥルーシブな手法がその精度の高さから一般的である。現在、音声強調処理/雑音抑圧処理の音声明瞭度客観評価指標として広く用いられているものに、STOI及びESTOIがある。なお、STOIについては、I.C.H. Taal, R.C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisyspeech,"

40

IEEE Tran. ASLP, vol.19, no.7, pp.2125-2136, 2011. [2]参照。ESTOIについては、J. Jensen and C.H. Taal, "An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers," IEEE/ACM Trans. ASLP, vol.24, no.1

50

1, pp.2009-2022, 2016.参照。

【 0 0 4 1 】

S T O I 及び E S T O I のような従来手法のほとんどが、評価信号と基準信号の音圧レベル差を無くす正規化処理を行っている。その後続く相関係数等の精度を高めるために重要である。しかしながら、このことにより、評価信号と基準信号の音圧が本質的に異なる補聴器信号に対する難聴者の音声了解度を予測することが困難となっている。また、個別の難聴者の音声了解度もオーディオグラム等の特性を反映できないため、対応不可能である。このため、S T O I 及び E S T O I のような従来の指標は、難聴者に対する適切な音声信号処理の開発に活用することは困難と考えられる。

【 0 0 4 2 】

従来手法の中では、例外的に、H A S P I では、評価信号と基準信号の間の音圧正規化処理を行っておらず、オーディオグラム等の特性を反映できる定式化がされている。H A S P I については、J.M. Kates and K.H. Arehart, "Coherence and the speech intelligibility index," JASA, vol.117, no.4 Pt 1, pp.2224-2237, 2005.参照。この文献では、補聴器信号処理の評価ができることと主張されている。しかし、その評価指標は、お互いに全く別の複数の評価尺度を組み合わせて算出している。それぞれの尺度を採用した明確な根拠の記述もなく場当たり的に見え、実際に使うとパラメータ設定に苦労する。少なくとも2～3個の測定結果からでは、パラメータは正確に決定できない。このため、補聴の対象となる特定の難聴者に個別対応することは、多数のデータ点をとる必要性から負担が軽い。

【 0 0 4 3 】

そこで、G E S I は、この点を解消し、難聴者個別に対応できる新しい音声了解度客観指標を提供することを目的としている。

【 0 0 4 4 】

< 2 . 2 G E S I >

【 0 0 4 5 】

図1は、本発明の音の評価装置の一例として、G E S I のための信号処理を行う装置10を示している。装置10は、音の基準信号(基準音声)と評価信号(評価音声)とを入力として受け付け、音の評価指標を計算する。音の評価指標は、装置10から出力され得る。実施形態において、基準信号と評価信号とが装置10に与えられる際に、両信号の信号レベルを揃える正規化処理は行われず。また、正規化処理の必要もない。したがって、基準信号と評価信号との間にレベル差があっても、両信号はそのまま装置10へ入力される。

【 0 0 4 6 】

装置10は、例えば、コンピュータプログラムを実行するコンピュータによって構成される。コンピュータは、プロセッサと、プロセッサに接続された記憶装置とを備える。記憶装置は、例えば、一次記憶装置及び二次記憶装置を備える。一次記憶装置は、例えば、R A M である。二次記憶装置は、例えば、ハードディスクドライブ(H D D) 又はソリッドステートドライブ(S S D) である。メモリは、プロセッサによって実行されるコンピュータプログラムを備える。プロセッサは、メモリに格納されたコンピュータプログラムを読み出して実行する。コンピュータプログラムは、G E S I の信号処理をコンピュータに実行させるためのプログラムコードを有する。コンピュータプログラムは、コンピュータを後述の分析器20及び評価器30として機能させるためのプログラムコードを有する。G E S I の信号処理の一部又は全部は、ハードウェアロジック回路によって実行されてもよい。

【 0 0 4 7 】

図1に示すように、装置10は、分析器20と、評価器30と、を備える。分析器20は、音の評価信号と基準信号とが入力信号として与えられ、両信号を分析する。分析器20は、入力信号の特徴量を求める。つまり、分析器20は、入力信号として与えられた評価信号と基準信号とを分析し、基準信号の第1特徴量と、評価信号の第2特徴量と、を出

10

20

30

40

50

力する。分析器 20 に与えられる両信号には、前述のように正規化処理は行われない。

【0048】

評価器 30 は、分析器 20 から出力された特徴量から、評価信号に対する評価データを生成する。ここでの、評価データは、一例として、音の客観評価指標である G E S I によって、評価信号を評価した値を示す。G E S I は、評価信号及び基準信号を、ガンマチャープ聴覚フィルタバンク (G C F B) と変調周波数フィルタバンク (M F B) との組み合わせで分析し、特徴量間の拡張コサイン類似度を統合して得られる指標である。

【0049】

なお、評価信号についての評価指標生成された評価データは、装置 10 の記憶装置に保存され得る。また、生成された評価データは、図示しない出力装置によって出力され得る。出力装置は、例えば、評価データを表示するディスプレイ、及び、表示データを外部へ送信する通信装置のいずれかである。

10

【0050】

図 1 に示す分析器 20 は、複数のフィルタバンク 11, 13 を備える。より具体的には、分析器 20 は、聴覚フィルタバンク 11 と、振幅包絡信号抽出部 12 と、変調フィルタバンク 13 と、を備える。

【0051】

聴覚フィルタバンク 11 は、聴取者の聴覚末梢系のモデルである。聴覚フィルタバンク 11 は、聴取者の受聴特性を示す。受聴特性は、例えば、難聴者特性である。信号が、聴覚フィルタバンク 11 に与えられると、信号は、受聴特性が反映されたものに変換される。つまり、聴覚フィルタバンク 11 は、入力信号を聴取者の聴こえ方に応じた信号に変換する。聴覚フィルタバンク 11 は、例えば、動的圧縮型ガンマチャープフィルタバンクである。

20

【0052】

聴覚フィルタバンク 11 は、 I (I は 2 以上の整数) 個のチャンネルを有する。聴覚フィルタバンク 11 は、 I 個のチャンネルの聴覚フィルタを備える。 I 個のチャンネルは、それぞれ、周波数帯域が異なる。聴覚フィルタバンク 11 は、入力信号それぞれを、 I 個のチャンネルそれぞれで周波数分析する。聴覚フィルタバンク 11 は、各チャンネルの聴覚フィルタを通過した信号を、そのチャンネルの帯域を通過した信号を、その帯域の応答の時間信号として出力する。聴覚フィルタバンク 11 は、各チャンネルにおいて、評価信号に対応する時間信号と、基準信号に対応する時間信号と、を出力する。

30

【0053】

振幅包絡信号抽出部 12 は、聴覚フィルタバンク 11 が出力した各時間信号の振幅包絡信号を抽出する。振幅包絡信号抽出部 12 は、例えば、入力された各時間信号に対して低域通過フィルタを適用して、振幅包絡信号を求める。振幅包絡信号抽出部 12 は、評価信号に対応する振幅包絡信号を、各チャンネルにおいて出力するとともに、基準信号に対応する振幅包絡信号を、各チャンネルにおいて出力する。

【0054】

変調フィルタバンク 13 は、変調周波数領域における複数の帯域通過フィルタを備えて構成され得る。変調フィルタバンク 13 は、 J (J は 2 以上の整数) 個のチャンネルを有する。変調フィルタバンク 13 は、 J 個のチャンネルのフィルタを備える。 J 個のチャンネルは、それぞれ周波数帯域が異なる。変調フィルタバンク 13 は、入力された各信号に対して、変調フィルタバンク 13 におけるチャンネル毎に、出力信号を計算し、出力する。例えば、聴覚フィルタバンク 11 のチャンネル数 I が 100 であり、変調フィルタバンク 13 のチャンネル数 J が 8 である場合、変調フィルタバンク 13 には、評価信号に対応する 100 個の信号が入力され、評価信号に対応する 800 個の信号が出力される。また、変調フィルタバンク 13 には、基準信号に対応する 100 個の信号が入力され、基準信号に対応する 800 個の信号が出力される。

40

【0055】

本実施形態において、変調フィルタバンク 13 の出力信号は、聴覚フィルタバンク 11

50

及び変調フィルタバンク 1 3 という複数のフィルタバンクによって周波数分析された信号である。本実施形態においては、一例として、変調フィルタバンク 1 3 の出力信号が、評価信号及び基準信号それぞれの特徴量として用いられる。

【 0 0 5 6 】

一例として、聴覚フィルタバンク 1 1 及び振幅包絡信号抽出部 1 2 は、図 2 に示す G C F B_{v23} (ガンマチャープフィルタバンク 1 1 A) によって構成され得る。G C F B_{v23} についての詳細は、入野俊夫, “ガンマチャープ聴覚フィルタに基づく模擬難聴システム whis の新実装,” The 3rd Japan-Taiwan Symposium on Psychological and Physiological Acoustics, 日本音響学会聴覚研究会資料, 第51 巻, pp.545-550, 2021. 参照。

【 0 0 5 7 】

G C F B_{v23} の出力は、フレームシフト 0 . 5 m s ごとの励起パターン (excitation pattern, EP) の時系列であり、絶対閾値に対しての相対レベルとして表現される聴覚的なスペクトログラムとなる。ここでの絶対閾値は、オーディオグラム上の値である聴力レベル (Hearing Level) から算出される値となっている。基準が明確に定義されていることから、入力された 2 つの信号の聴覚特徴量である E P を直接比較できる。G C F B_{v23} では、フレームシフト 0 . 5 m s ごと (サンプリング周波数 2 0 0 0 H z) に正值の EP が出力されるので、包絡線も自動的に求まっている。

【 0 0 5 8 】

変調フィルタバンク 1 3 は、一例として、図 2 に示すように、I I R フィルタで構成され変調フィルタバンク 1 3 A であり得る。G C F B_{v23} の各チャンネルの出力信号は、I I R フィルタで構成された変調フィルタバンク 1 3 A で分析される。変調フィルタバンク 1 3 A は、一例として、カットオフ周波数 1 H z の低域通過フィルタと、8 周波数 (2 H z , 4 H z , 8 H z , 1 6 H z , 3 2 H z , 6 4 H z , 1 2 8 H z , 2 5 6 H z) の中心周波数を持つ帯域通過フィルタと、を備える。ガンマチャープフィルタバンク 1 1 A から出力された信号は、変調周波数ごとの分析波形として、変調フィルタバンク 1 3 A から出力される。

【 0 0 5 9 】

図 1 に戻り、評価器 3 0 は、変調フィルタバンク 1 3 の出力信号を、評価信号及び基準信号の特徴量として用いて、評価信号の評価指標を計算する。図 1 に示す評価器 3 0 は、一例として、類似度計算部 1 4 と、重み付け和計算部 1 5 と、音評価指標計算部 1 6 (評価データ生成部 1 6) と、を備える。また、評価器 3 0 は、パラメータ調整部 1 7 と、基本周波数抽出部 1 8 と、を備える。

【 0 0 6 0 】

図 2 は、類似度計算部 1 4、重み付け和計算部 1 5、及び、音評価指標計算部 1 6 それぞれにおける計算方法の例を示している。

【 0 0 6 1 】

類似度計算部 1 4 は、評価信号の第 1 特徴量 $m^t_{ij}()$ と、基準信号の第 2 特徴量 $m^r_{ij}()$ とから、類似度 S_{ij} を計算する。類似度 S_{ij} は、図 2 に示す式 (1) に従って計算され得る。類似度 S_{ij} は、分析器 2 0 における周波数分析のチャンネル $i j$ ごとに計算される。ここでのチャンネル数は $I \times J$ である。ここで、 i は、G C F B_{v23} のチャンネルを示し、 i は 1 から I までの整数である。 j は、変調フィルタバンク 1 3 A のチャンネルを示し、 j は 1 から J までの整数である。 t は、フレーム時刻を示す。G E S I では、この類似度 S_{ij} が、客観評価指標の基礎とされる。

【 0 0 6 2 】

$m^t_{ij}()$ は、G C F B_{v23} のチャンネル i 、および、変調フィルタバンク 1 3 のチャンネル j のフレーム時刻における出力であって、評価信号に対応するものを示す。また、 $m^r_{ij}()$ は、G C F B_{v23} のチャンネル i 、および、変調フィルタバンク 1 3 のチャンネル j のフレーム時刻における出力であって、基準信号に対応するものを示す。

【 0 0 6 3 】

図 2 に示す式 (1) では、拡張コサイン類似度が計算される。拡張コサイン類似度は、

10

20

30

40

50

通常のコサイン類似度を、後述のように調整可能パラメータを用いて拡張したものである。

【0064】

図2に示す式(1)では、特徴量 m_{ij}^t ()、 m_{ij}^r ()は、フレーム時間方向に統合するため、時間 の区間内において加算される。 の加算区間は、例えば、評価信号及び基準信号となる音声の長さに対応する時間である。例えば、音声が1つの単語を発音したものであれば、 の加算区間は、その単語を発音したときの時間長さに相当する。なお、音声が、文を発話した音声である場合のように時間的に長い信号である場合、基準信号において振幅が小さい部分で切り分けて分析されてもよい。この場合、 の加算区間は、切り分けられ部分の時間長さに相当してもよい。

10

【0065】

式(1)において、パラメータ は、評価信号と基準信号の音圧レベル差などに対応するための重み値である。式(1)において、 方向に統合された第1特徴量 m_{ij}^t ()及び第2特徴量 m_{ij}^r ()それぞれは、パラメータ によって重み付けられている。ここでは、パラメータによって重み付けられた第1特徴量を「重み付き第1特徴量」といい、同じく調整パラメータによって重み付けられた第2特徴量を「重み付き第2特徴量」という。式(1)においては、重み付き第1特徴量と重み付き第2特徴量とを用いて、類似度 S_{ij} が計算される(式(1)の分母参照)。なお、式(1)においては、重み付けされていない第1特徴量と第2特徴量も用いられる(式(1)の分子参照)。

20

【0066】

式(1)において、パラメータ は、0以上1以下の数である。 の値が0.5であるとき、式(1)は、通常のコサイン類似度を計算する式になる。式(1)において、 は、固定値ではなく、可変値である。パラメータ は、パラメータ調整部17によって調整される。パラメータ調整部17は、パラメータ の大きさの指定する入力をユーザから受け付けるように構成されていてもよいし、所定のアルゴリズムに基づいてパラメータ の大きさを計算するように構成されていてもよい。パラメータ調整部17によって設定されたパラメータ は、類似度計算部14に与えられる。パラメータ の値を調整することで、類似度 S_{ij} を計算する上で、第1特徴量と第2特徴量との重み付けを変えることができる。

30

【0067】

通常のコサイン類似度は、2つのベクトルの内積を、2つのベクトルのノルムの積で除した値である。第1特徴量 m_{ij}^t ()及び第2特徴量 m_{ij}^r ()についての通常のコサイン類似度を求める場合、第1特徴量 m_{ij}^t ()及び第2特徴量 m_{ij}^r ()の内積を、第1特徴量 m_{ij}^t ()のノルム及び第2特徴量 m_{ij}^r ()のノルムの積で除した値となる。第1特徴量 m_{ij}^t ()のノルムは、第1特徴量 m_{ij}^t ()の二乗の総和を、 $1/2$ 乗して求められる。また、第2特徴量 m_{ij}^r ()のノルムは、第2特徴量 m_{ij}^r ()の二乗の総和を、 $1/2$ 乗して求められる。これに対して、実施形態に係る拡張コサイン類似度 S_{ij} を求める際には、第1特徴量 m_{ij}^t ()のノルムは、第1特徴量 m_{ij}^t ()の二乗の総和を、 α 乗して求められる。また、第2特徴量 m_{ij}^r ()のノルムは、第2特徴量 m_{ij}^r ()の二乗の総和を、 $(1 - \alpha)$ 乗して求められる。つまり、実施形態における類似度は、通常のコサイン類似度を α によって重み付けた値(重み付け特徴量)を、ノルムとして扱って計算されるコサイン類似度になっている。

40

【0068】

例えば、評価信号が基準信号に対して音圧レベルに差がある場合など、評価信号の音響特性が基準信号に対して異なる場合がある。この場合、基準信号に対する評価信号の音圧レベルなどの評価信号音響特性に応じて を調整することで、両信号の音響特性の違いに応じた重みを、両信号の特徴量を与えて、類似度 S_{ij} を適切に計算することができる。

【0069】

また、聴取者にとっての音声了解度は、音の音響特性だけでなく、聴取者の受聴特性又

50

は受聴環境特性によっても影響を受ける。さらに、聴取者にとっての音声了解度は、音響特性、受聴特性、及び受聴環境特性のうちの複数の特性の組み合わせによって影響を受けることもある。したがって、音を評価するにあたっては、聴取者の受聴特性又は受聴環境特性を考慮すべき場合がある。したがって、 α は、これらの様々な特性を考慮して、調整されてもよい。 α の値の設定の仕方の例については後述される。聴取環境における閾値上レベルに関連する値に設定すると予測がよくなる。

【 0 0 7 0 】

続いて、各チャンネル i, j の類似度 S_{ij} が、全体の類似度指標 d_s へ統合される。図 2 の式 (2) に示すように、類似度指標 d_s は、全ての類似度 S_{ij} の重み付け和として計算される。式 (2) において、 w_i は、 i 方向の重みであり、 w_j は、 j 方向の重みである。 w_i 及び w_j は関数であってもよいし、定数であってもよい。以下においては、一例として、 w_i は関数とし、 w_j は、定数とする。 w_i 及び w_j は、例えば、音声知覚の知見に基づいて、装置 1 0 の設計時に適宜決定される。

10

【 0 0 7 1 】

本発明者の検討の結果、G C F B チャンネル方向の重み関数 w_i を、すべて 1 とすると、後述の < 2 . 4 > の結果を説明できないことが判明した。これは、低い周波数の EP は、主に声帯振動の影響が大きく、音素識別にそれほど有効でないためと考えられる。

【 0 0 7 2 】

ここでは、声帯振動の影響を EP から効果的に取り除くことができる、Size-Shape Image (S S I) 重みを使用した。S S I 重みについては、T. Matsui, T. Irino, R. Uemura, K. Yamamoto, H. Kawahara, and R.D. Patterson, "Modelling speaker-size discrimination with voiced and unvoiced speech sounds based on the effect of spectral lift," Speech Communication, vol.136, pp.23-41, 2022.参照。

20

【 0 0 7 3 】

S S I 重みは、音声からの話者寸法推定の聴覚計算理論である安定化 Wavelet-Mellin 変換の枠組みにおける知覚対応表現を、聴覚的スペクトログラム上で近似できるようにしたものである。安定化 Wavelet-Mellin 変換については、T. Irino and R.D. Patterson, "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-mellin transform," Speech Communication, vol.36, no.3-4, pp.181-203, 2002.参照。

30

【 0 0 7 4 】

ここでは、図 2 に示す式 (3) で、重み関数 w_i を決定した。式 (3) において、 $f_{p,i}$ は、G C F B における i 番目のチャンネルのフィルタのピーク周波数であり、 h_{max} は、上限値を決定するパラメータである。また、式 (3) の重み関数 w_i は、評価信号の基本周波数をパラメータとして有する。評価信号の基本周波数は、基本周波数抽出部 1 8 によって評価信号から求められ、重み付け和計算部 1 5 に与えられる。重み付け和計算部 1 5 は、当たられた基本周波数によって決定される重み関数 w_i と、定数である重み w_j と、を用いて、全チャンネル i, j の類似度 S_{ij} を統合した類似度指標 d_s を計算する。

【 0 0 7 5 】

音評価指標計算部 1 6 は、類似度指標 d_s を、音の客観的指標の一例である音声了解度 I (単語正解率 %) に変換する。この変換のための計算式は、例えば、図 2 に示す式 (4) である。式 (4) は、シグモイド関数である。式 (4) において、 a, b は定数であり、例えば、後述の主観評価実験の一部の結果 (たとえば、強調音声処理無しの結果) から決められる。

40

【 0 0 7 6 】

式 (4) によって計算された音声了解度を示すデータ (評価データ) は、図示しない出力部を介して、装置 1 0 から出力され得る。

【 0 0 7 7 】

< 2 . 3 主観評価実験 >

【 0 0 7 8 】

50

以下、GESIの評価のために用いた、音声了解度主観評価実験について説明する。この実験の手続きは、入野俊夫，田丸萌夏，山本絢子，“模擬難聴システムWHISの新実装と末梢系特性の音声了解度への影響，”音講論(春)，no.2-4-13，pp.665-668，2022．に記載の実験手続と同様である。また、以下では、クラウドソーシング遠隔実験の結果も示す。

【0079】

< 2.3.1 WHISを用いた音声了解度実験 >

【0080】

模擬難聴システムWHISを用いて、難聴者の聞こえに相当する聞き取りにくい音声を作り、それを健聴者に聞かせてどの程度の音声了解度となるかを調べた。これは、認知系の機能低下等の影響を排除して、聴覚末梢系の機能低下の影響だけを評価することを目的としていた。なお、WHISについては、入野俊夫，“ガンマチャープ聴覚フィルタに基づく模擬難聴システムWHISの新実装，”The 3rd Japan-Taiwan Symposium on Psychological and Physiological Acoustics，日本音響学会聴覚研究会資料，第51巻，pp.545-550，2021．参照。

10

【0081】

< 2.3.1.1 音声条件 >

【0082】

実験では、親密度別単語了解度試験用音声データセット(FW07)に収録されている男性話者(mis)の最低親密度4モーラ単語音声を使用した。FW07については、近藤久久，天野成昭，坂本修一，鈴木陽一，“親密度別単語了解度試験用音声データセット2007 (fw07)，”音声資源コンソーシアム，国立情報学研究所，2007．参照。

20

【0083】

この音声に対しバブル雑音を重畳し、信号対雑音比(SNR)を-3dBから9dBの間で3dBごとに設定した。これを原音声(Unprocessed、以下、“Unpro”という)とした。模擬難聴条件では、圧縮特性健全度 = 0.5と設定し、70歳男性(以下、“70yr”という)と80歳(以下、“80yr”という)の平均聴力レベルに設定してWHISで模擬音を合成した。70歳男性の平均聴力レベルは、ISO 7029:2017 Acoustics - Statistical distribution of hearing thresholds related to age and gender参照。80歳の平均聴力レベルは、立木孝，笹森史朗，南吉昇，一戸孝七，村井和夫，村井盛子，河嶋寛，“日本人聴力の加齢変化の研究，”Audiology Japan，vol.45，no.3，pp.241-250，2002参照。

30

【0084】

さらに、高域の聴力レベル低下の影響を明確化するため、原音声を単純-20dB音圧低減(以下、“-20dB”という)した条件を加えた。これは、80歳模擬音とほぼ同じL_qレベルとなっている。

【0085】

< 2.3.1.2 実験条件 >

【0086】

4種類の信号処理と5つのSNR条件の組み合わせの全20条件の各条件に対し、FW07の1組20単語を割り当て、単語総数は400(=20×4×5)となった。これを、Webベースの実験環境[17]で音声提示と回答収集を行なった。Webベースの実験環境については、A. Yamamoto and et.al., “Comparison of remote experiments using crowdsourcing and laboratory experiments on speech intelligibility,” Proc. Interspeech 2021, pp.181-185, 2021.参照。

40

【0087】

全部聴取するには約2時間かかるため、2つのタスクに分割した。実験は、防音室での実験と、クラウドソーシングサービスを用いた遠隔実験の2種類を行なった。実験は倫理委員会の承認を受けている。

【0088】

< 2.3.1.2.1 防音室実験 >

50

【0089】

まず、防音室において聴取実験をおこなった。暗騒音レベルは約 $26 \text{ dB} (L_{Aeq})$ で、パソコンから DAC (SONY, NW-A55) を通しヘッドホン (SONY, MDR-1AM2) で聴取した。音圧は L_{eq} で、 65 dB とした。参加者は、19歳から24歳の日本人大学生13名である。音声了解度実験に初参加で、 125 Hz から 8000 Hz の範囲で聴力レベル 20 dB 以下と健聴な聴力レベルであることを確認した。

【0090】

< 2.3.1.2.2 クラウドソーシング遠隔実験 >

【0091】

クラウドソーシングサービス (Lancers 社) を用いて、遠隔実験を行った。音声了解度実験に初めて参加する者を募った。最初のタスクでは先着順で参加者を募り、2番目のタスクでは第1タスクを完全に終了した人に直接依頼した。参加者は、静かな部屋で有線ヘッドホン/イヤホンで聴取することが指示された。音圧レベルは“Unpro”の場合に快適に聞けるレベルに調整させ、“ -20 dB ”の時にはなんとか聞き取れるレベルとなっていることを確認させた。最終的に、28名の日本人参加者が2タスクとも完遂した。年齢は22歳から66歳と幅広く、3名が難聴傾向に関する質問肢に「はい」と回答していた。

10

【0092】

< 2.3.1.3 トーンピップテスト >

【0093】

遠隔実験においては、参加者の聴力レベル、設定した音圧レベル、再生機器の特性、暗騒音レベル等の聴取条件はまったくわからない。そこで少しでも聴取条件を推定するために、トーンピップ (短時間正弦波) を用いた検査を導入した。トーンピップの詳細は、A. Yamamoto and et.al., “Subjective intelligibility of speech sounds enhanced by ideal ratio mask via crowdsourced remote experiments with effective data screening,” Submitted to Interspeech 2022 arXiv:2203.16760 参照。

20

【0094】

トーンピップテストでは、参加者に -5 dB ずつ減少する15個の短時間正弦波を聞かせ、各周波数で何個 (N_{pip}) 聞こえたかを回答させた。トーンピップの周波数は音声帯域をカバーできるように、 500 Hz , 1000 Hz , 2000 Hz , 4000 Hz とした。この結果から、“Unpro”音声の閾値上レベルを推定できる。トーンピップテストの結果は、参加者の聴取環境 (受聴環境特性) や聴力レベル (受聴特性) を反映している。

30

【0095】

< 2.3.2 実験結果 >

【0096】

実験参加者の回答データから、音声了解度を4モーラ単語の完全正答率として求めた。図3は防音室実験の結果を示し、図4は遠隔実験の結果を示す。図3, 4において、縦軸は、正解率であり、横軸はSNRである。図3, 4は、それぞれ音声了解度の参加者平均と標準偏差とを示している。防音室実験・遠隔実験の両方において、SNR向上により正答率も向上し、傾向も似ていることがわかる。ただし、遠隔実験の方が被験者間のばらつきが大きい。防音室実験においては、“Unpro”と“ -20 dB ”との間で、音圧差が大きいかも関わらず了解度への影響がほとんどないことがわかる。これに対し、遠隔実験では、了解度が大きく異なっている。同じ音声を聴取しているにもかかわらず、異なるのは、聴取条件の違いが反映されているものと考えられる。“70yr”の了解度はこれらより低く、“80yr”ではさらに下がっている傾向は、防音室実験・遠隔実験の両方において同様である。

40

【0097】

“ -20 dB ”に対して“80yr”は、音声波形での音圧レベル L_{eq} は同じだが、高域のレベルが異なる。これにより、子音を含む高域成分が聞き取りにくくなったため了解度が低下したと考えられ、従来知見と整合性がある。

【0098】

50

< 2.4 客観評価結果 >

【0099】

客観評価指標は、防音室実験結果である図3と、遠隔実験結果である図4の両方とも同時に説明できることが望まれる。以下では、本発明の実施例であるGESIが、図3, 4の両方とも説明できる良好な客観評価指標であることを示す。

【0100】

< 2.4.1 実施例：GESIによる予測 >

【0101】

防音室実験・遠隔実験の両方の結果の相違は聴取条件の違いに起因するものであり、聴取条件の違いが反映され得るトーンピップ聴取数 N_{pip} で、その違いに対応できる。

10

【0102】

そこで、GESIにおける類似度の式(1)内のパラメータの値を、トーンピップ聴取数 N_{pip} に基づいて求めた。パラメータの値は、例えば、以下の式(5)によって求められ得る。

$$= 0.50 + 0.02 \cdot (15 - N_{pip}) \quad (5)$$

【0103】

実験の全参加者平均のトーンピップ聴取数 N_{pip} は、防音室実験において12.5個であり、遠隔実験で10.0個であった。これらのトーンピップ聴取数 N_{pip} を式(5)に代入すると、防音室実験の場合、 $= 0.55$ であり、遠隔実験の場合、 $= 0.60$ となる。これらのパラメータ値を式(1)に代入して、GESIの予測結果を計算した。なおここで、参加者が聞いた音声のうち20単語分を用いて“Unpro”の音声了解度と2乗誤差が最小となるように、式(4)のa, bを決定した。

20

【0104】

図5は、 $= 0.55$ の場合の予測結果を示し、図6は、 $= 0.60$ の場合の予測結果を示す。図5は図3に、図6は図4によく似ていることがわかる。特に、図5, 6は、図3, 4における“Unpro”と“-20 dB”の位置関係をよく再現できている。このことは、パラメータによって位置関係を制御できることを示している。また、図5, 6と図3, 4との間で、“70yr”、“80yr”の相対位置関係も良くあっている。したがって、式(5)を使ってパラメータを求めれば、個人ごとの心理物理曲線のある程度良く予測できる。

30

【0105】

WHISでは、GCFBへの末梢系の機能低下度合いパラメータ設定に対応して、精度良く模擬難聴音が合成できていることがわかっている。GESIも同じGCFBを基盤としているため、難聴者個人ごとの末梢系起因の音声了解度低下を良く再現できる。

【0106】

< 2.4.2 比較例：従来法ESTOI, STOIによる予測 >

【0107】

図7は、ESTOIによる予測結果である。ESTOIは多くの従来客観手法と同様、評価信号と基準信号のレベル差を正規化した上で評価指標を計算する。したがって、原理的に“Unpro”と“-20 dB”は全く同じとなる。図7に示すように、今回の条件では“70yr”、“80yr”も一致してしまい、さらにばらつきも極めて大きかった。すなわち、防音室実験及び遠隔実験の結果を全く説明できないことがわかる。なお、図示は省略するが、STOIによる予測結果も、ESTOIによる予測結果と同様の結果になった。

40

【0108】

< 2.4.3 比較例：従来法HASPIによる予測 >

【0109】

図8は、HASPIによる予測結果である。HASPIは、従来客観手法の中では例外的に、評価信号と基準信号のレベル差を正規化しない。この利点が働いて、図8に示すように、図3(防音室実験結果)に近い予測値を出せた。しかし、HASPIは、図4(遠隔実験結果)を全く説明することができなかった。HASPIにおいて、図4を説明する

50

ために必要なパラメータは具体的に提案されておらず、図 4 を説明するための調整は難しい。さらに H A S P I では、‘ 70yr ’、‘ 80yr ’ の予測値は実験値よりもかなり低い。また、H A S P I では複数の独立の特徴量をシグモイド関数の中で統合して了解度を算出する。このため、‘ Unpro ’ の 2 ~ 3 個の実験値だけでは、H A S P I のためのパラメータを精度良く決定できない。これらのことから、H A S P I では、難聴者個人ごとの特性を少ない負担で予測することは難しい。

【 0 1 1 0 】

< 2 . 5 G E S I についてのまとめ >

【 0 1 1 1 】

本開示は、新しい客観評価指標 G E S I を提案し、アルゴリズムを説明した。また、模擬難聴システム W H I S を用いた主観評価実験を述べ、防音室実験とクラウドソーシング遠隔実験において、全く同じ刺激音に対して音声了解度の違いが出たことが示された。この違いは、例えば、聴取条件における閾値上レベルにより生じていると考えられる。さらに、これらの音声了解度を、G S E I では良く予測できるのに対し、従来法の S T O I , E S T O I , H A S P I では予測できないことが示された。G E S I を使うことにより、音声強調を含む補聴処理音を聴取させた場合の難聴者個人ごとの音声明瞭度を、予測できるようになる。

10

【 0 1 1 2 】

本発明は、上記実施形態に限定されるものではなく、様々な変形が可能である。

【 符号の説明 】

20

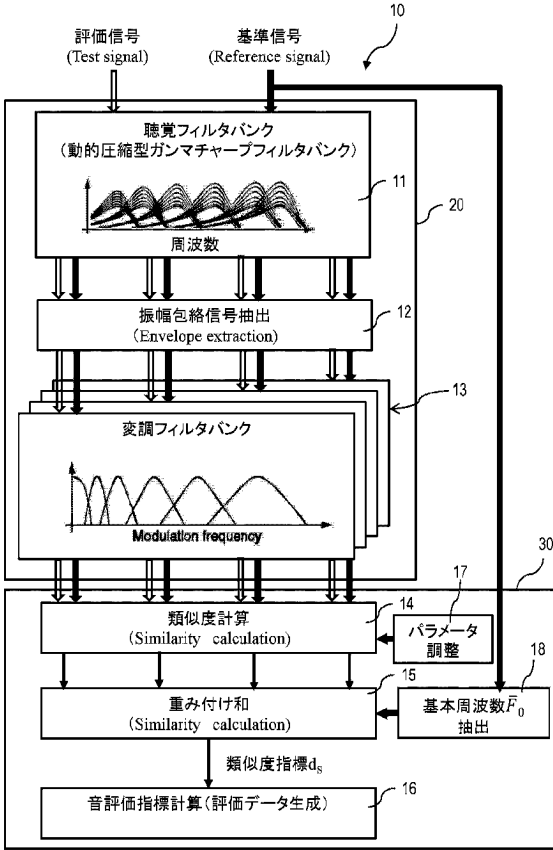
【 0 1 1 3 】

- 1 0 : 評価装置
- 1 1 : 聴覚フィルタバンク
- 1 1 A : ガンマチャープフィルタバンク
- 1 2 : 振幅包絡信号抽出部
- 1 3 : 変調フィルタバンク
- 1 3 A : 変調フィルタバンク
- 1 4 : 類似度計算部
- 1 5 : 重み付け和計算部
- 1 6 : 音評価指標計算部 (評価データ生成部)
- 1 7 : パラメータ調整部
- 1 8 : 基本周波数抽出部
- 2 0 : 分析器
- 3 0 : 評価器

30

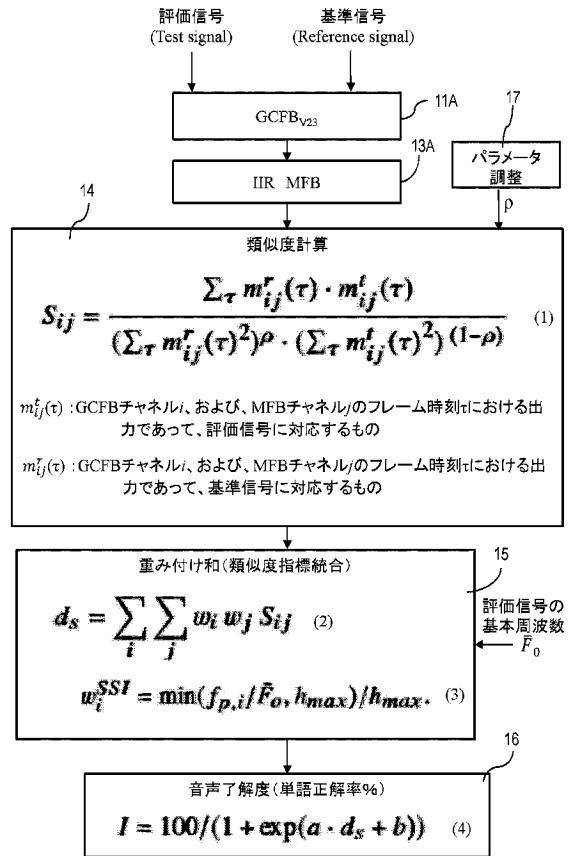
【図1】

図1



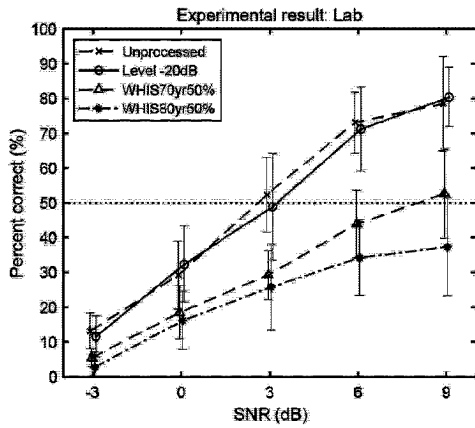
【図2】

図2



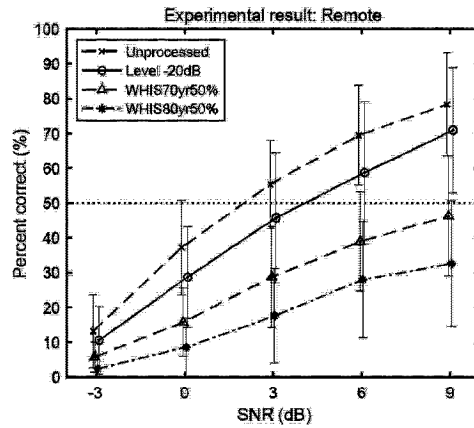
【図3】

図3



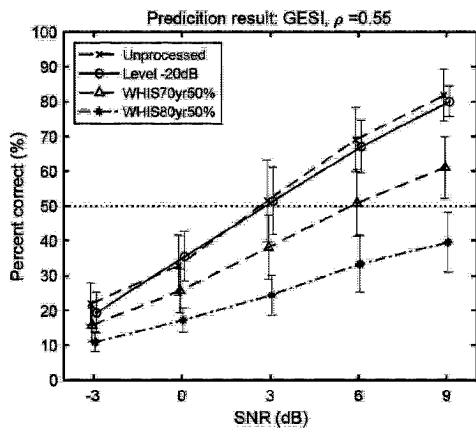
【図4】

図4



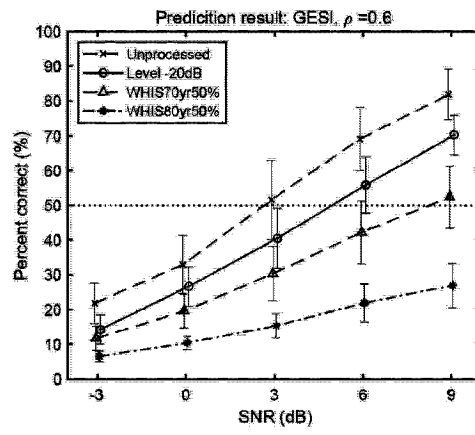
【 図 5 】

図5



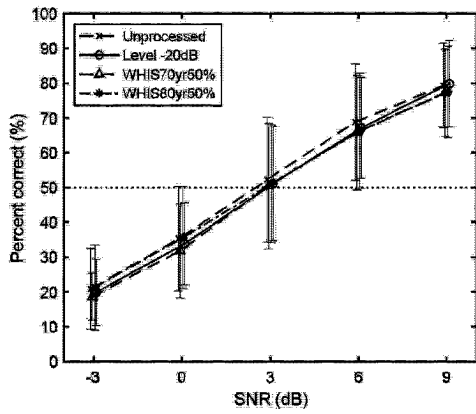
【 図 6 】

図6



【 図 7 】

図7



【 図 8 】

図8

